

ПРЕГЛЕДНИ ЧЛАНАК

УДК 004.6/.7:027:004.43XML

ПРЕГЛЕД РАЗВОЈА ДИГИТАЛНИХ БИБЛИОТЕКА БАЗИРАНИХ НА XML ТЕХНОЛОГИЈАМА

Мирјана Зеремски
Природно-математички факултет, Нови Сад

Сажетак

У области библиотекарства најпознатији су UNIMARC и MARC формат за опис библиографске грађе и формирање библиографских записа. Развој библиотечких информационих система базираних на овим стандардима све више користе XML технологије, као водеће технологије на пољу чувања и размене информација у електронском облику. У овом раду дат је преглед значајнијих пројеката развоја дигиталних библиотека базираних на библиотечким стандардима и коришћењу XML технологија за моделирање и имплементацију система.

Кључне речи: дигиталне библиотеке, XML технологије, библиографски стандарди

SURVEY ON DEVELOPMENT OF DIGITAL LIBRARIES BASED ON XML TECHNOLOGIES

Mirjana Zeremski
Faculty of Science and mathematics, Novi Sad

Abstract

In librarianship field the most popular standards for describing of bibliographic units and forming of bibliographic records are MARC and UNIMARC. In the development of librarian information systems, based on these standards, XML technology is used more and more, as one of leading technologies in a field of storing and exchanging electronic information. The survey of significant projects on digital libraries which are based on bibliographic standards and usage of XML technologies for modeling and implementing the system is presented in this paper.

Key words: digital libraries, XML technologies, bibliographic standards

1. УВОД

Истраживање у области дигиталних библиотека је у свету веома актуелно у последњих 10 година. Једна дефиниција концепта *дигитална библиотека* се може наћи на сајту [1], а гласи овако: "Концепт *дигитална библиотека* није просто еквивалентан колекцији дигитализованих података са алатом за управљање информацијама. То је пре окружење које треба да обједини колекције, сервисе и људе који би подржали пуни животни циклус који чини креирање, ширење, употреба и чување података, информација и знања." Наравно, поред ове постоје и друге дефиниције. Тако је у раду [2] дат преглед концепата на основу којих се дефинише дигитална библиотека као и концепата који су у непосредној вези са дигиталним библиотекама. У овом раду разматран је само део који се односи на развој дигиталних библиотека са информатичког аспекта, посебно дела који се односи на XML технологије.

Наведена дефиниција јасно указује на мултидисциплинарни карактер дигиталних библиотека. У овом раду разматран је само део који се односи на развој софтверских система за дигиталне библиотеке, посебно дела који се односи на примену XML технологија у тим системима. У наредном параграфу дат је кратак преглед истраживања у области дигиталних библиотека, затим су наведене значајније научне конференције и часописи који су посвећени тој области. У преосталом делу рада детаљније су приказани пројекти доступни преко Интернета који су у свом развоју користили XML технологије.

2. ДИГИТАЛНЕ БИБЛИОТЕКЕ

На основу истраживања спроведеног од стране удружења IEEE (Institute of Electrical and Electronics Engineers), односно његовог Техничког комитета за дигиталне библиотеке (Technical Committee on Digital

Libraries - TCDL) [3] основаног 1997. године, кључни аспекти за које је потребно дати решења приликом развоја дигиталних библиотека су:

- чување података,
- кориснички интерфејс,
- класификација и индексирање,
- претраживање информација,
- испоручивање садржаја,
- презентација података,
- администрација и конзервација.

Интензивно истраживање из области дигиталних библиотека у САД је 1995. године довело до формирања Федерације дигиталних библиотека (Digital Library Federation - DLF) од стране шеснаест институција. Циљ ове организације је унапређење употребе електронско – информационих технологија са циљем проширења њихових колекција и сервиса. Истраживања спроведена од стране ове Федерације приказана су кроз чланке објављене у периоду од 1999. до 2002. године, на њеном сајту [4]. Овде се као један од кључних аспеката за успешну имплементацију и употребу дигиталне библиотеке наводи појам организације библиотеке, за чију реализацију је потребно користити један или више система за организацију знања (Knowledge Organization System - KOS). Још један кључни аспект који је овде разматран је чување дигиталних информација. Услед брзог напретка у развоју хардвера и софтвера, тренутно је тешко гарантовати дуговечност и читљивост дигиталних информација само у једној људској генерацији. Да би се превазишао овај проблем потребно је обезбедити ефикасне методе за приступ информацијама креираним помоћу застарелог софтвера и хардвера, или обезбедити трансфер дигиталних материјала из једне софтверско/хардверске конфигурације у другу. Следећи проблем везан за развој дигиталних библиотека се односи на то да се у таквом једном online систему мења улога библиотеке, као и њени односи са корисницима или другим информационим сервисима.

Да је развој дигиталних библиотека веома актуелан и да брзо напредује види се и по мноштву научних конференција чија је основна тема развој дигиталних библиотека, а које се одржавају широм света [5]. На пример:

- У Европи се, почев од 1997. године, сваке године одржава *Европска конференција о истраживањима и напредним технологијама у области дигиталних библиотека* (ECDL - European Conference on Research and Advanced Technology for Digital Libraries).
- Почев од 1999. године у Русији се сваке године одржава *All-Russian Scientific Conference on Digital Libraries: Advanced Methods And Technologies, Digital Collections*.
- Међународној конференцији *International Conference on Enterprise Information Systems - ICEIS*, која се одржава у Европи, је од 2001. године придружена и редовна пратећа радионица везана за дигиталне библиотеке - *Workshop on New Developments in Digital Libraries*.
- У САД се редовно одржава *The Joint Conference on Digital Libraries*.
- У Азији се од 1998. године одржава *ICADL - International Conference of Asian Digital Libraries*.
- Поред наведених редовних годишњих конференција, сваке године се широм света одржи и низ појединачних конференција на ову тему.

Поред конференција које прате развој дигиталних библиотека, постоје и многи часописи из исте области [5]. На пример:

- *Ariadne Magazine* намењен библиотекарима британских академских библиотека,
- *Journal of Digital Information - JoDI* у издању *Oxford University Press*,
- *International Journal on Digital Libraries* издавача *Springer-Verlag*,
- *D-lib Magazine* чији је издавач *Corporation for National Research Initiatives*,
- *The National Digital Library Program* који издаје *Library of Congress*,
- *International Technology and Libraries - ITAL* који издаје *Library and Information Technology Association*,
- *Russian Digital Libraries Journal* - руски часопис о дигиталним библиотекама, као и многи други.

На основу свега наведеног, види се да развој дигиталних библиотека подразумева истраживања из многобројних рачунарских области: вештачке интелигенције, база података, мрежног програмирања, претраживања и презентације мултимедијалних података, компресије различитих формата мултимедијалних података, програмирања корисничког интерфејса, итд. Потреба и интересовање за даљим развојем ове области су веома изражени у целом свету.

У раду [2] је дат кратак историјски преглед развоја дигиталних библиотека, при чему је наглашено основно тежиште сваког од пројеката који су наведени. У овом раду су приказани неки јавно доступни пројекти у којима је употребљена XML технологија за моделирање MARC (MAchine Readable Cataloguing) [6] или UNIMARC (UNiversal MAchine Readable Cataloguing) [7] стандарда. То су:

- XMLMARC пројекат Станфорд универзитета,
- MARC XML пројекат Конгресне библиотеке САД,
- Пројекат Националне медицинске библиотеке САД,
- BiblioML пројекат под покровитељством Министарства културе и комуникација Републике Француске.
- Библиотечки софтверски систем БИСИС развијен на Универзитету у Новом Саду

3. XMLMARC ПРОЈЕКАТ

Развој пројекта XMLMARC је започет 1998. године на Станфорд универзитету, Станфорд, Калифорнија, за потребе његове медицинске библиотеке – Lane Medical Library [8]. Окосница овог пројекта је Java клијент/сервер апликација која врши конверзију MARC записа у XML документе.

У циљу испитивања софтвера за конверзију и могућности за креирање извештаја и омогућавање приступа корисницима путем Интернета, у Lane Medical Library је конвертовано преко 200000 постојећих библиографских и UNIMARC/Authorities записа у XML документе. На тај начин су добијени библиотечки подаци који се лако могу интегрисати са другим Web ресурсима.

Следећи задатак у развоју овог система је био испитивање могућности складиштења и претраживања XML библиографских записа [9]. Испитане су четири могућности:

- релациона база података,
- објектно-оријентисана база података,
- native XML база података,
- могућност чувања записа у фајл систему, уз коришћење софтвера за претраживање фајл система са XML документима.

Стручњаци са овог пројекта дају следећи закључак везан за складиштење XML записа: после овако детаљно урађеног испитивања они ће и даље наставити да складиште XML записе у Oracle бази података у CLOB колонама, али ће такође пратити даљи развој XML native база података.

Током 2002. године у раду на XMLMARC пројекту објављен је нови резултат од изузетног значаја [10], [11], [12]. XOBIS (The XML Organic Bibliographic Information Schema) је шема написана у RELAX NG шема језику, која представља реализацију идеје под називом *organic bibliography* први пут поменути у штампаном облику 2000. године. Ова идеја подразумева да се хијерархијске везе које постоје у библиографским записима, превазиђу коришћењем скупа темпоралних веза (проузрокованих променом наслова, имена, извештаја, итд.). Такве везе су дефинисане и за записе нормативне контроле (authority). На овај начин је добијена Web - оријентисана мрежа међусобних веза заснована на патернима који се стално понављају. Уочено је да се исти патерни понављају и у везама између библиографских записа и authority записа, као и у међусобним везама различитих врста записа.

4. MARC XML ПРОЈЕКАТ

Library of Congress (Конгресна библиотека) САД и служба за развој MARC стандарда развијају систем за рад са MARC подацима у XML окружењу [13], [14]. Овај систем обухвата многе компоненте као што су: шеме, stylesheet-ови и софтверски алати које развија и одржава Конгресна библиотека.

Сви суштински подаци из MARC записа су конвертовани и изражени у XML документу. Тачније, само подаци везани за физичку структуру MARC записа нису пренети у XML документ. Тако је практично добијена конверзија без губитака, а реализован је и обрнути смер конверзије - из XML документа у MARC запис.

Поред описане шеме, у оквиру овог пројекта је развијена још једна шема названа Metadata Object Description Schema, скраћено MODS, чија намена је, по наводу аутора, да буде носилац изабраних података из постојећих MARC записа, као и да омогући креирање оригиналних записа [15]. Она садржи подскуп MARC поља и користи описна имена тагова (тј. имена чије значење одговара садржају тага) уместо нумеричких.

Архитектура MARC XML система састоји се из три нивоа. Дакле, основу система чине MARC записи и алат који врши конверзију тих записа у XML записе. Да би ово било могуће, на следећем нивоу је дефинисана XML шема која описује XML библиографске записе. На основу ове шеме могуће је на трећем нивоу развијати софтверске алате за обраду XML MARC записа који се могу поделити у три категорије:

- за трансформацију, што подразумева конверзију између XML MARC записа и других формата за чување мета података (као што је MODS),
- за презентацију, што подразумева приказивање MARC података у читљивој форми,
- за анализу, што подразумева обраду MARC података у циљу добијања аналитичког резултата, као што је валидација.

James (Java MARC Events) је Java API (Application Programming Interface) за обраду MARC записа коришћен на овом пројекту [16].

Поред модела догађаја, James садржи и објектни модел MARC записа који омогућава конверзију из MARC формата у разне друге формате (као што је на пример конверзија у XML документе), а уз помоћ модела догађаја. Важно је нагласити да ова конверзија подразумева и конверзију кодног распореда ANSEL који се користи у MARC формату у UCS/Unicode кодни распоред. James такође врши и валидацију MARC XML записа спрема шеме написане у RELAX NG шема језику. За валидацију се користе пакети Jing и Sun Multi-Schema Validator - MSV.

5. ПРОЈЕКАТ НАЦИОНАЛНЕ МЕДИЦИНСКЕ БИБЛИОТЕКЕ САД

У новембру 1999. године Национална медицинска библиотека САД (U.S. National Library of Medicine) је објавила да приступа завршној фази транзиције до новог система за креирање и управљање подацима. У првој фази је било предвиђено да се нови систем испроба на MEDLINE библиотеци. Дотадашњи систем ELHILL је чувао податке у сопственом ELHILL Unit Record Format – EURF формату, а у новом систему записе у том формату треба да замене XML документи.

Током 2000. године аутори овог пројекта су обезбедили да постојећа размена записа почне да се одвија посредством XML докумената преко ftp сервера.

На жалост, на сајту [17] не постоји опис, документација, нити било какава информација о софтверу помоћу којег се врши конверзија из EURF записа у XML документе, као ни податак о томе да ли су у развоју овог пројекта предузети даљи кораци после 2000. године.

6. BiblioML ПРОЈЕКАТ

Овај пројекат се развија од 1999. године под покровитељством Министарства за културу и комуникацију Републике Француске. То је XML апликација за UNIMARC библиографске записе и UNIMARC/Authorities записе.

Дата су два DTD-а: један за библиографске, а други за UNIMARC/Authorities записе. Предвиђено је преузимање записа из постојеће UNIMARC базе, и њихово трансформисање у XML документе. Апликација која врши конверзију је реализована у Java окружењу, и састоји се од biblioml.jar пакета развијеног у оквиру овог пројекта, и SAX парсера. Оба DTD-а, као и пакет biblioml.jar су доступни на сајту [18].

Последњи подаци везани за развој овог пројекта су ажурирани 2000. године.

7. БИБЛИОТЕЧКИ СОФТВЕРСКИ СИСТЕМ БИСИС

Развој система БИСИС је започет 1993. године на Универзитету у Новом Саду [19]. У верзији 3.0 развијен је сопствени текст сервер за индексирање и претраживање библиографских записа у UNIMARC формату. Подршка за Unicode стандард доследно је спроведена у целокупном систему БИСИС верзија 3.0.

У оквиру даљег развоја БИСИС-а предвиђено је да се у реализацији његових функција, где је то могуће, укључи коришћење XML технологије. У ту сврху је показано да је могуће извршити моделирање UNIMARC формата и библиографских записа по овом формату помоћу XML Schema језика [20], [21]. Такође је обезбеђено да XML библиографски записи, добијени конверзијом од библиографских записа по UNIMARC формату, буду коректни [22].

У претходно описаним пројектима, као и у прегледу доступне литературе, није разматрано моделирање UNIMARC формата. У магистарској тези [23] је главни циљ био да се анализира могућност моделирања комплетног UNIMARC формата помоћу XML Schema језика. Пошто је YUMARC [24] подскуп UNIMARC формата у [23] је показано да се комплетан YUMARC формат може описати помоћу XML Schema језика. Пошто су сви концепти UNIMARC формата заступљени у YUMARC формату, то значи да се и UNIMARC формат може у потпуности моделирати XML Schema језиком.

За разлику од осталих разматраних пројеката, у којима је главни циљ био обезбедити XML записе који ће се користити у размени и презентацији података, у [23] је разматрана могућност моделирања библиографских стандарда тако да се у потпуности опишу помоћу XML Schema језика. Показано је да је записе по UNIMARC формату, односно све концепте који се користе у записима, могуће у потпуности моделирати помоћу XML Schema језика. Моделирање је извршено тако да називи тагова у XML библиографском запису одговарају ознакама поља и потпоља из UNIMARC записа. У осталим разматраним пројектима примењени су другачији приступи: или су таговима дата описна имена која указују на врсту података садржану у том елементу (при томе касније остаје нејасно из ког поља, тј. потпоља у UNIMARC/MARC запису је преузет тај податак), или је комбиновано коришћење описних имена и имена која садрже идентификаторе поља или потпоља, или су идентификатори поља и потпоља наведени као вредности атрибута елемената који одговарају пољима, тј. потпољима.

На основу добијених резултата у [23], у току је израда магистарске тезе [25] под називом *XML едитор за опис UNIMARC формата*. У Java програмском окружењу дизајниран је едитор који омогућава ажурирање XML документа UNIMARC формата, при чему кориснику није потребно познавање XML-а. Такође је омогућена и валидација тако креираног документа, а на основу шеме UNIMARC формата дате у [23].

8. ЗАКЉУЧАК

XML технологија је присутна у развоју библиотечких информационих система. То је и било за очекивање зато што су библиографски стандарди структурирани, те као такви су веома погодни за моделирање и имплементацију у XML технологији. Могући правци даљих истраживања у овој области су:

- Развој специјализованих едитора стандарда за опис библиографске грађе.
- Преузимање библиографских података са специјализованих сервера базираних на XML документима.
- Контрола XML библиографских записа.
- Развој специјализованог едитора за креирање и ажурирање XML библиографских записа.
- Стварање кооперативних библиотечких информационих система.
- Развој сервера за узајамну каталогизацију базираног на XML технологији.

ЛИТЕРАТУРА:

- [1] Santa Fe Workshop on Distributed Knowledge Work Environments,
<http://www.si.umich.edu/SantaFe>
- [2] Крстев, Ц., *Дигиталне библиотеке – разграничење појмова*, ИНФОтека, 2002, год. 3, бр. 1-2, стр. 3-11
- [3] IEEE Technical Committee on Digital Libraries (TCDL),
<http://www.ieee-tcdl.org>
- [4] Digital Library Federation. Home Page,
www.diglib.org
- [5] Gross, B., *Digital Library Information and Resources*, 2002.
<http://www.canis.uiuc.edu/~bgross/dl>
- [6] MARC STANDARDS, Library of Congress - Network Development and MARC Standards Office
<http://www.loc.gov/marc/>
- [7] UNIMARC Manual: bibliographic format / International Federation of Library Association and Institutions, IFLA Universal Bibliographic Control and International MARC Programme, New Providence, London, 1994.
- [8] Medlane XMLMARC, A Project of Lane Medical Library, Stanford University Medical Center
<http://xmlmarc.stanford.edu/MLA2001/medlane.html>
- [9] Clarke, K., *Medlane/XMLMARC Update: From MARC to XML Database*, MLA 2001.
<http://xmlmarc.stanford.edu/MLA2001/medlane.html>
- [10] Miller, Dick R. *Bibliographic access management at Lane Medical Library: fin demillennium experimentation and bruised-edge innovation*. Cataloging & Classification Quarterly 2000; 30(2/3):139-66. Cf. p. 164.
<http://xmlmarc.stanford.edu/dick.pdf>
- [11] Miller, Dick R., *Adding Luster to Librarianship: XML as an Enabling Technology*, MLGSCA/NCNMLG Joint Meeting, Scottsdale, 2002
<http://elane.stanford.edu/laneauth/Luster.html>
- [12] Miller, Dick R., Clarke, Kevin S., *XOBIS: The XML Organic Bibliographic Information Schema*, Lane Medical Library, Stanford University, Apr.-Sept. 2002.
<http://elane.stanford.edu/laneauth/XOBIS.pdf>
- [13] MARC XML – MARC 21 XML Schema, Official Web Site
<http://www.loc.gov/standards/marcxml/>
- [14] *Implementation Guidelines for the Open Archive Initiative Protocol for Metadata Harvesting – An XML Schema to represent MARC records*, Protocol version 2.0 of 2002-06-14.
http://www.openarchives.org/OAI/2.0/guidelines-oai_marc.htm
- [15] MODS – Metadata Object Description Schema, Official Web Site
<http://www.loc.gov/mods>
- [16] James (Java MARC Events)
<http://www.bpeters.com>
- [17] *New Data Format and Distribution Media for Licensees of NLM Data*, U.S. National Library of Medicine, November 19, 1999.
http://www.nlm.nih.gov/bsd/xml_announce.html
- [18] *BiblioML Project*
<http://www.culture.fr/BiblioML/en/index.html>

- [19] Surla, D., Konjović, Z., Milosavljević, B., Vidaković, M., Bibliotečki informacioni sistem ver. 3.01, IX Међународна конференција "Informatika u obrazovanju, kvalitet i nove informacione tehnologije", Zrenjanin, 2000. str. 494-504
- [20] Zereński, M., Surla, D., Modeliranje UNIMARC formata pomoću XML Schema jezika, XXVIII Jugoslovenski simpozijum o operacionim istraživanjima SYM-OP-IS 2001, Beograd, 2001., str. 261-264.
- [21] Zereński, M., Surla, D., Modeliranje bibliografskih zapisa pomoću XML Schema jezika, XXVIII Jugoslovenski simpozijum o operacionim istraživanjima SYM-OP-IS 2001, Beograd, 2001., str. 265-268.
- [22] Zereński, M., Surla, D., Validacija bibliografskih zapisa pomoću XML Schema jezika, *CD-ROM Simpozijuma o računarskim naukama i informacionim tehnologijama YU INFO 2002*, Kopaonik, 2002.
- [23] Зеремски, М., *Моделирање UNIMARC формата у XML технологији*, магистарска теза, Нови Сад, 2002.
- [24] COMARC/B Format za bibliografske podatke - COBISS, Univerzitetni institut informacijskih znanosti IZUM, Maribor, 1992.
- [25] Мијић, В., *XML едитор за опис UNIMARC формата*, прихваћена тема магистарске тезе на Природно-математичком факултету у Новом Саду.